Advances in direct methods for protein crystallography

Isabel Usón and George M Sheldrick*

Recent advances in *ab initio* direct methods have enabled the solution of crystal structures of small proteins from native X-ray data alone, that is, without the use of fragments of known structure or the need to prepare heavy-atom or selenomethionine derivatives, provided that the data are available to atomic resolution. These methods are also proving to be useful for locating the selenium atoms or other anomalous scatterers in the multiple wavelength anomalous diffraction phasing of larger proteins at lower resolution.

Addresses
Lehrstuhl für Strukturchemie, Universität Göttingen, Tammannstraße 4, D37077 Göttingen, Germany
*e-mail: gsheldr@shelx.uni-ac.gwdg.de

Current Opinion in Structural Biology 1999, **9**:643–648

0959-440X

Abbreviations
MAD multiple wavelength anomalous diffraction
MIR multiple isomorphous replacement

Introduction

In an X-ray diffraction experiment, the intensities of the

calculation of an electron density map, which provides an interpretable picture of a molecule, requires both intensities and phases. The standard methods of solving this problem, known as the 'crystallographic phase problem', for macromolecules involve either using a closely related structure as a search fragment (molecular replacement) or preparing isomorphous heavy-atom derivatives (multiple isomorphous replacement, MIR). The latter method, although of enormous historical significance, often requires considerable patience and skill. It can be particularly frustrating when no related structure is available for molecular replacement (sometimes even an NMR-derived structure of the same protein does not work) and when no suitable heavy-atom derivative can be prepared. An alternative approach that is gaining rapidly in popularity is MAD (multiple wavelength anomalous diffraction) phasing; at the cost of extra time spent in the wet lab expressing and crystallising a protein with its methionine residues replaced by selenomethionine and with very careful synchrotron measurements of the reflection intensities at three or more different X-ray wavelengths, an electron density map can be obtained in a more routine manner. Even MAD phasing involves a hidden phase problem, however, as it can prove to be difficult to find the selenium atoms when there are more than about 20 crystallographically independent sites and without these atoms to provide reference phases, no map can be calculated.

Small-molecule structures of up to about 100 unique atoms (not including hydrogen) are routinely solved using so-called 'direct methods', without the use of such subterfuges; all that is required is a single set of native diffraction intensities, which is needed anyway for the refinement of the structure. These direct methods use sophisticated probability theory and the assumption of approximately equal, resolved atoms to estimate reflection phases from the measured intensities. This is a way of exploiting the fact that a crystal structure is over-determined at atomic resolution, that is, there are many more measured intensities than parameters that are necessary to describe an atomic model. Until very recently, only a handful of unknown structures with over 200 unique atoms had been solved using direct methods, so extending them to solving the structures of typical macromolecules with thousands of atoms did not look promising.

A crucial breakthrough came in 1993, with the introduction of dual-space iteration, also known as 'Shake-and-Bake', by the Buffalo group (in particular Weeks, Miller, Hauptman and colleagues) [1]. This overall strategy has been implemented in two computer programs, SnB (Miller, Weeks and co-workers) [2,3••,4•,5••,6] and SHELXD (Sheldrick)

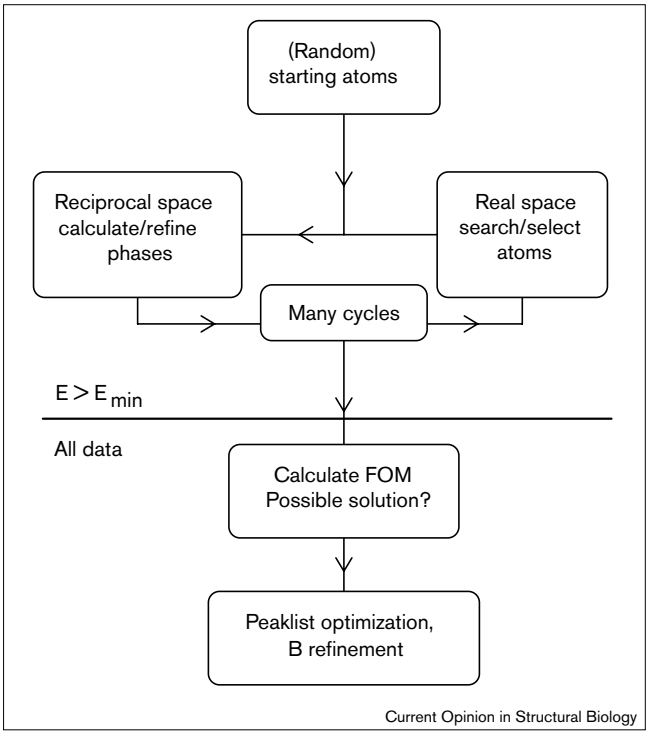
the publication of the review by Hauptman [9] in this journal. For a full review of 'conventional' direct methods, that is, up to the Shake-and-Bake revolution, we recommend the chapter by Giacovazzo [10]. We should mention that a full account of dual-space methods, giving much more mathematical detail than is possible here, will appear soon [11]. An alternative approach, based on Karle–Hauptman determinants [12], that is also able to solve larger structures than conventional direct methods was implemented in the program CRUNCH by De Graaff and co-workers [13••] and has been reviewed recently in this journal [14•].

In keeping with recent practice, we will use the term 'direct methods' to refer to methods for solving the phase problem using probability theory and '*ab initio*' for methods that employ native data only, without the use of phase information from isomorphous derivatives or from anomalous scattering. *Ab initio* methods may exploit general features of protein structure, such as the presence of solvent or disulfide bridges, but may not use specific structural information in the form of a search fragment from a related structure. Thus, the use of the Shake-and-Bake procedure in the location of selenium atoms from MAD data is an application of direct methods, but not *ab initio* methods.

Dual-space recycling methods

Figure 1 shows a scheme of the procedure that is common to dual-space recycling methods. Each trial starts

Figure 1



The general scheme of dual-space recycling. E_{min} is the minimum normalised structure factor E used in the calculations. FOM, figure of merit.

with a new set of random atoms. It is possible to constrain the starting atoms in various ways: the type and number of atoms; consistency with the Patterson function (which is calculated from the measured reflection intensities without the need for phase information) [15]; interatomic distance restrictions and so on. Phases are calculated from the starting atoms and different strategies can be applied in reciprocal space to try to improve these phases; this half of the procedure is referred to as 'shaking'. The two strategies most frequently applied are minimisation of the minimal function and using the tangent formula.

The minimal function [1,16] is the weighted mean square difference between the current and statistically expected cosines of the sums of three phases. These phase triplets must be chosen such that their reflection indices add up to zero in reciprocal space. Any minimisation algorithm could be used, but, in practice, the parameter shift method [17] has the advantage of simplicity and gives good results; each phase may be shifted once or twice by a given value and the shifted phases are adopted if they lead to a reduction in the minimal function. Usually, phase shifts of 90° are employed, but sometimes — especially for space group P1 — larger shifts are required [6]. This is the default algorithm in the SnB program.

The tangent formula [18] forms the basis of most conventional direct methods programs, either to refine all the phases or in a Karle-type tangent expansion [19] to derive phases for the remaining reflections from the phases that are best determined by the current atoms. The latter is the default in SHELXD.

A map is then calculated using the improved phases, combined with normalised structure factors (E values) derived from the observed intensities and searched for the highest peaks that are considered to be possible atoms. In this real-space half of the procedure ('baking'), the following three strategies have been most widely employed:

1. Simply select the highest N peaks as atoms, where N is the expected number of unique nonhydrogen atoms, usually ignoring the solvent. This is the current default in SnB. Optionally, the atoms may be assigned different element types, for example, for a protein, the highest peaks might be assigned as sulfur, followed by oxygen and so on.

2. Peaklist optimisation [15]. The list of N peaks is scanned twice from bottom to top, testing each atom in turn. The atom is eliminated if this leads to the improvement of a figure of merit, such as the correlation coefficient between observed and calculated normalised structure factors (E_{obs} and E_{calc}). Although this method works, it is inefficient for large structures, because it requires more computer time than the other two methods.

3. Random omit maps (GM Sheldrick, unpublished data). A given percentage — say 30% — of the top N peaks are deleted at random. This turns out to be a surprisingly effective search algorithm when combined with tangent expansion and it also requires the least computer time. This is the default in SHELXD.

The remaining atoms are then used to calculate new phases and start a new cycle. The optimum number of cycles per trial is structure-dependent, but a value of about N/2 is usually suitable [3••]. Dual-space recycling is always performed using only the largest normalised structure factors (E values) (about 10–20% of the total data), firstly because the minimal function and tangent formula are only valid for large E values and secondly because it results in a substantial saving in computer time. Even so, the procedure is computationally expensive and would not be practicable without the recent improvement in computer performance.

A figure of merit is needed to identify possible solutions and also hopeless trials (ideally, at an early stage), so that they can be either pursued or discarded and the next trial started with the generation of a new set of random starting atoms. Suitable figures of merit are the minimal function (provided that it is calculated with the phases derived from the atoms, before phase refinement), the crystallographic R factor between E_{obs} and E_{calc}, or the correlation coefficient based on all the data [20]. Only promising trials are

Table 1

Some structures solved using dual-space methods.

Compound	Space group	N(mol)†	N(+solv)‡	N(heavy)§	Resolution (Å)#	Program used**	References††
Vancomycin	$P4_32_12$	202	312	6Cl	0.9–1.4	S, D	[32,33]
Actinomycin X2	P1	273	305	–	0.90	D	
Actinomycin Z3	$P2_12_12_1$	186	307	2Cl	0.96	D	[34]
Actinomycin D	P1	270	314	–	0.94	D	[34]
Gramicidin A*	$P2_12_12_1$	272	317	–	0.86–1.1	S, D	
DMSO d6 peptide	P1	320	326	–	1.20	S	
Er-1 pheromone	C2	303	328	7S	1.00	S	[35]
Ristocetin A	$P2_1$	294	420	–	1.03	D	
Crambin*	$P2_1$	327	423	6S	0.83–1.2	S, D	[36]
Hirustasin	$P4_32_12$	402	467	10S	1.2–1.55	D	[25]
Cyclodex. deriv.	$P2_1$	448	467	–	0.88	D	[37]
Alpha-1 peptide	P1	408	471	Cl	0.92	S	[38]
Rubredoxin*	$P2_1$	395	497	Fe, 6S	1.0–1.1	S, D	
Vancomycin	P1	404	547	12Cl	0.97	S	[39]
BPTI*	$P2_12_12_1$	453	561	7S	1.08	D	
Cyclodex. deriv.	$P2_1$	504	562	28S	1.00	D	
Balhimycin*	$P2_1$	408	598	8Cl	0.96	D	
Mg–complex*	$P\bar{1}$	576	608	8Mg	0.87	D	
Scorpion toxin II*	$P2_12_12_1$	508	624	8S	0.96–1.2	S, D	[40]
Amylose CA26	P1	624	771	–	1.10	D	[41]
Mersacidin	$P3_2$	750	826	24S	1.04	D	[42]
Cv HiPIP H42Q*	$P2_12_12_1$	631	837	4Fe	0.93	D	[43]
HEW lysozyme*	P1	1001	1295	10S	0.85	S, D	[23]
rc-WT Cv HiPIP	$P2_12_12_1$	1264	1599	8Fe	1.20	D	[43]
Cytochrome c_3	$P3_1$	2024	2208	8Fe	1.20	D	[24]

*Previously known structures (prior to the advent of Shake-and-Bake). †N(mol) is the number of unique nonhydrogen atoms in the protein or other molecules being investigated. ‡N(+solv) is N(mol) plus the number of unique nonhydrogen atoms in the solvent. §N(heavy) is the number and type of the atoms heavier than oxygen. #When two numbers are given, the second indicates the lowest resolution at which truncated data have yielded a solution. **The program codes are SnB (S) and SHELXD (D). ††Reference numbers are only given when the details of the structure solution by dual-space methods are given in the paper.

expanded using all data through E-Fourier recycling without phase refinement (SnB), or peaklist optimisation (SHELXD) to get as complete a solution as possible. This solution may be improved further by refinement of the isotropic B values (thermal displacement parameters) of the atoms in the unexpurgated solution. This yields a map that is easier to interpret using standard protein-chain-tracing methods. Alternatively, the program wARP [14•,21,22] may be used to complete the solution.

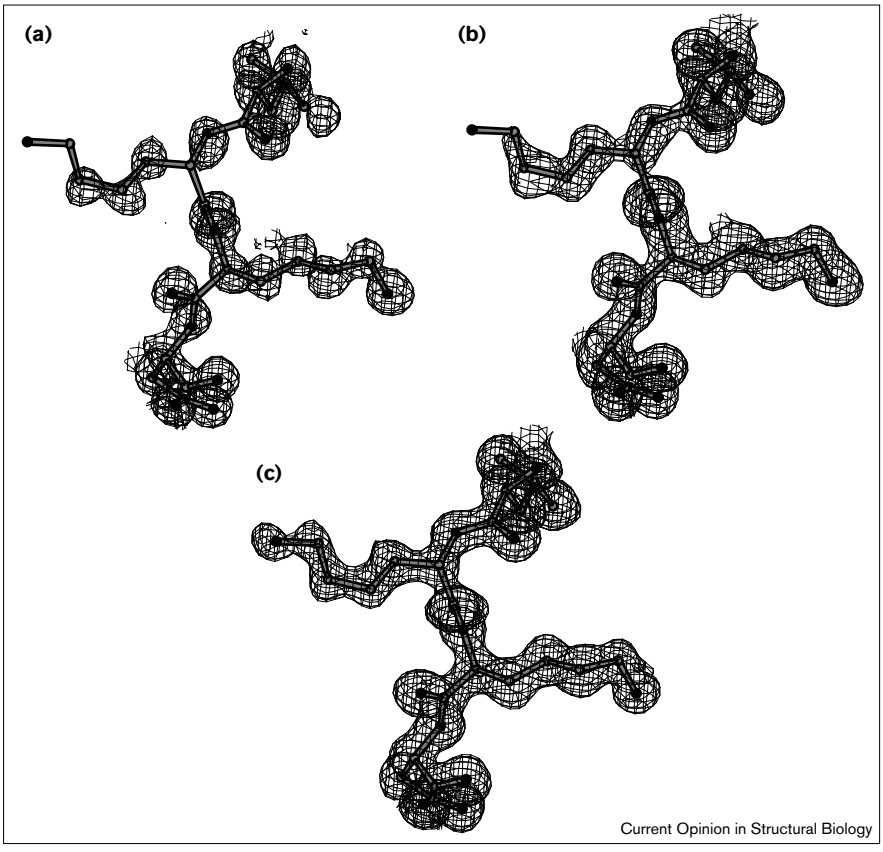
Recent successes of dual-space recycling methods

Table 1 provides a list of structures with more than 250 independent atoms that have been solved using Shake-and-Bake-based methods. The list is incomplete because some unpublished structures have been left out at the request of those involved. Some previously unknown structures (prior to the advent of Shake-and-Bake) had defeated conventional methods despite exhaustive attempts. So far, the largest structure solved that contains no atom heavier than sulfur is triclinic lysozyme with 1001 protein atoms, first solved using SHELXD and, shortly afterwards, using SnB [23]. The largest unknown structure with no atom heavier than sulfur is the peptide antibiotic mersacidin; although there are only 20 amino acids in the molecule, it crystallises with six independent molecules, so that its diffraction properties are similar to those of a 120-residue protein. The presence of heavier elements, such as iron in haem groups or iron–sulfur clusters, allows larger structures to be solved, the record so far being 2024 protein atoms in a cytochrome c_3 [24]. It is to be expected that structures with more atoms will become amenable to direct methods as computers become more powerful. The resolution limit seems to be a tougher barrier and, as structures get bigger, they tend to diffract to lower resolution anyway. Most of the structures that were solved diffracted to 1.1 Å or better; however, this limit can be relaxed somewhat if a few heavy atoms are present, especially if they possess the lowest B values in the structure, as is often the case.

An example of the solution of a small protein

The 55 amino acid protein hirustasin [25] provides an example of a successful *ab initio* solution after molecular replacement and heavy-atom methods had failed, even though an isomorphous gold derivative had been prepared. It subsequently also proved to be possible to solve the structure independently using the gold derivative and the program SHARP [26], despite difficulties arising from the pseudosymmetric gold position. The SHELXD program was able to solve the structure using either of the two native data sets (1.2 Å low temperature and 1.4 Å room temperature). The structure, which contains five disulfide

Figure 2



Electron density maps of the hirustasin structure, showing part of the mainchain and two lysine sidechains. **(a)** 1.4 Å *ab initio* map, **(b)** 1.4 Å sigma-A-weighted *ab initio* map after the refinement of B values and **(c)** 1.2 Å sigma-A-weighted $2mF_o–DF_c$ map after structure refinement. Note that in the original publication [25] of the 1.2 Å maps, the figures corresponding to (a) and (b) were accidentally exchanged. The *ab initio* map is artificially 'atomic' but, after B refinement, the map already closely resembles that calculated from the final model, although no structural information has been assumed.

bridges, could even be solved with data truncated to 1.55 Å. Figure 2 shows the final model in the area of residues 37–40 and maps at different stages of the structure determination, based on the 1.4 Å room temperature data. The map in Figure 2a, calculated with normalised structure factors and the *ab initio* phases, shows the location of the majority of the atoms in the structure. Easier to trace, although less sharp, is the sigma-A-weighted map [27] in Figure 2b, calculated after isotropic B-value refinement of all of the atoms in the unexpurgated *ab initio* solution. The quality of the *ab initio* maps is striking if one bears in mind that they are generated using an automatic objective computational procedure from the native data alone; no assumptions about protein structure have been made, so they are entirely free of 'model bias'. The second, 1.2 Å low temperature map is very similar to the sigma-A-weighted $2mF_o–DF_c$ map [27] after structure refinement that is shown in Figure 2c, the correlation between the two being 0.81 and the mean phase difference being 26°.

Finding the selenium atoms in multiple wavelength anomalous diffraction phasing

Locating the positions of the selenium atoms or other anomalous scatterers is an essential step in the MAD approach to the phase problem [28•]. Both conventional small-molecule direct methods programs and automated Patterson function have been successfully employed to find the selenium atoms, but both methods run into problems for larger proteins when the number of selenium atoms exceeds about 20 (about one residue in 50 is methionine, so this corresponds approximately to a 1000 amino acid protein). Conventional direct methods require relatively complete data and are easily upset by individual aberrant reflection intensities; MAD data tend to be noisy and incomplete. When the anomalous scatterers are located using purely Patterson methods, each successive atom found is subject to coordinate errors that accumulate and degrade the solution when many atoms need to be located. It seems that dual-space direct methods are rather effective for solving such problems and both SnB and SHELXD have been successfully applied to structures with 20 to 70 independent seleniums. Most of these structures were solved using data to about 3 Å resolution, although it appears to be possible to find at least some of the anomalous scatterers at lower resolution. The Patterson function is still very useful both in the generation of starting atoms and in identifying the correct selenium atoms in the structure [29]. The location of the sulfur atoms and chloride ions in triclinic lysozyme using SHELXD from the single-wavelength (1.5 Å) anomalous differences [30] suggests that it may not always be necessary to incorporate selenium!

Conclusions

Dual-space direct methods are capable of solving small protein structures (up to approximately 1000 independent

protein atoms) and other macromolecular structures of similar size, provided that native data are available to atomic resolution, which, in practice, usually means 1.2 Å or better (hirustasin was not typical in this respect). The size of the structure is less important than the resolution of the data and, as computers become faster, size should present even less of a problem.

The wARP procedure [14•,21,22] currently requires approximately correct starting phases, but it then provides automated model extension involving the iterative location of potential atoms in difference maps and their refinement against the native data; poorly defined atoms are rejected and the cycle repeated. This procedure appears to give good results down to about 2.3 Å resolution, so an integration of the Shake-and-Bake and wARP algorithms has the potential to solve structures from the native data alone at more normal protein resolutions, albeit at enormous computational cost. Another promising, but computationally expensive, approach to reducing dependence on atomic resolution would be to search for groups of atoms (e.g. a piece of α helix), rather than for individual atoms, in the real-space stage [31]. Recent improvements in synchrotron beam-lines, cryocrystallography and area detectors make it easier to collect data to high resolution, so we anticipate that it will become increasingly common to solve protein structures using *ab initio* direct methods, even though these methods may never reach the omnipotence that they have achieved in small-molecule structure determination.

Acknowledgements

We are grateful to many SHELX users for providing test data and suggestions, to HA Hauptman, R Miller, CM Weeks, M Schäfer and TR Schneider for useful discussions about dual-space methods and to the Fonds der Chemischen Industrie for support.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Miller R, DeTitta GT, Jones R, Langs DA, Weeks CM, Hauptman HA: **On the application of the minimal principle to solve unknown structures.** *Science* 1993, **259**:1430-1433.

2. Miller R, Gallo SM, Khalak HG, Weeks CM: **SnB: crystal structure determination via Shake-and-Bake.** *J Appl Crystallogr* 1994, **27**:613-621.

3. •• Miller R, Weeks CM: **Shake-and-Bake: applications and advances.** In *Direct Methods for Solving Macromolecular Structures.* Edited by Fortier S. Dordrecht: Kluwer Academic Publishers; 1998:389-400.
An authoritative description of the Shake-and-Bake philosophy.

4. • Weeks CM, Miller R, Hauptman HA: **Extending the resolving power of Shake-and-Bake.** In *Direct Methods for Solving Macromolecular Structures.* Edited by Fortier S. Dordrecht: Kluwer Academic Publishers; 1998:463-468.
An investigation into the resolution that is required for successful structure solution using the Shake-and-Bake procedure.

5. •• Weeks CM, Miller R: **The design and implementation of SnB v2.0.** *J Appl Crystallogr* 1998, **32**:120-124.
A description of the latest version of the SnB computer program, now distributed.

6. Weeks CM, Miller R: **Optimizing Shake-and-Bake for proteins.** *Acta Crystallogr D* 1999, **55**:492-500.

7. Sheldrick GM: **Direct methods based on real/reciprocal space iteration.** In *Recent Advances in Phasing: Proceedings of the CCP4 Study Weekend (DL-CONF-97-001).* Edited by Wilson KS, Davies G, Ashton AS, Bailey S. Daresbury Laboratory, Warrington, UK: CCLRC; 1997:147-158.

8. • Sheldrick GM: **SHELX: applications to macromolecules.** In *Direct Methods for Solving Macromolecular Structures.* Edited by Fortier S. Dordrecht: Kluwer Academic Publishers; 1998:401-411.
A detailed description of the 'Half-Baked' alternative to SnB (also known as SHELXD).

9. Hauptman HA: **Phasing methods for protein crystallography.** *Curr Opin Struct Biol* 1997, **7**:672-680.

10. Giacovazzo C: **Direct methods.** In *International Tables for Crystallography,* vol B. Edited by Shmueli U. Dordrecht: Kluwer Academic Publishers; 1996:201-229.

11. Sheldrick GM, Hauptman HA, Weeks CM, Miller R, Usón I: **Direct methods: *ab initio* phasing.** In *International Tables for Crystallography,* vol F. Edited by Arnold E, Rossmann M. Dordrecht: Kluwer Academic Publishers; 1999:in press.

12. Karle J, Hauptman HA: **The phases and magnitudes of structure factors.** *Acta Crystallogr* 1950, **3**:181-187.

13. •• Van der Plas JL, De Graaff RAG, Schenk H: **Karle-Hauptman matrices and eigenvalues: a practical approach.** *Acta Crystallogr A* 1998, **54**:267-272.
The authors describe an entirely different approach to Shake-and-Bake, based on Karle–Hauptman determinants and incorporated in the program CRUNCH, that is more effective than conventional direct methods for problem structures.

14. • Abrahams JP, De Graaff RAG: **New developments in phase refinement.** *Curr Opin Struct Biol* 1998, **8**:601-605.
A review of the state-of-play in density modification and direct methods, published a year ago, with further details of the determinant approach.

15. Sheldrick GM, Gould RO: **Structure solution by iterative peaklist optimization and tangent expansion in space group P1.** *Acta Crystallogr B* 1995, **51**:423-431.

16. Debaerdemaeker T, Woolfson MM: **On the application of phase relationships to complex structures XXII. Techniques for random phase refinement.** *Acta Crystallogr A* 1983, **39**:193-196.

17. Bhuiya AK, Stanley E: **The refinement of atomic parameters by direct calculation of the minimum residual.** *Acta Crystallogr* 1963, **16**:981-984.

18. Karle J, Hauptman H: **A theory of phase determination for the four types of non-centrosymmetric space groups 1P222, 2P22, $3P_12$, $3P_22$.** *Acta Crystallogr* 1956, **9**:635-651.

19. Karle J: **Partial structural information combined with the tangent formula for noncentrosymmetric crystals.** *Acta Crystallogr B* 1968, **24**:182-186.

20. Fujinaga M, Read RJ: **Experiences with a new translation-function program.** *J Appl Crystallogr* 1987, **20**:517-521.

21. Perrakis A, Sixma TA, Wilson KS, Lamzin VS: **wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models.** *Acta Crystallogr D* 1997, **53**:448-455.

22. Perrakis A, Morris R, Lamzin VS: **Automated protein model building combined with interactive structure refinement.** *Nat Struct Biol* 1999, **6**:458-463.

23. Deacon AM, Weeks CM, Miller R, Ealick SE: **The Shake-and-Bake structure determination of triclinic lysozyme.** *Proc Natl Acad Sci USA* 1998, **95**:9284-9289.

24. Frazão C, Sieker L, Sheldrick GM, Lamzin VS, LeGall J, Carrondo MA: ***Ab initio* structure solution of a dimeric cytochrome c_3 from *Desulfovibrio gigas* containing disulfide bridges.** *J Biol Inorg Chem* 1999, **4**:162-165.

25. Usón I, Sheldrick GM, De La Fortelle E, Bricogne G, Di Marco S, Priestle JP, Grütter MG, Mittl PRE: **The 1.2 Å crystal structure of hirustasin reveals the intrinsic flexibility of a family of highly disulphide bridged inhibitors.** *Structure* 1999, **7**:55-63.

26. De La Fortelle E, Bricogne G: **Maximum-likelihood heavy-atom refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods.** *Methods Enzymol* 1997, **276**:472-494.

27. Read RJ: **Improved Fourier coefficients for maps using phases from partial structures with errors.** *Acta Crystallogr A* 1986, **42**:140-149.

28. Smith JL: **Multiwavelength anomalous diffraction in**
• **macromolecular crystallography.** In *Direct Methods for Solving Macromolecular Structures.* Edited by Fortier S. Dordrecht: Kluwer Academic Publishers; 1998:211-225.
An excellent account of MAD (multiple wavelength anomalous diffraction) phasing from a leading practitioner.

29. Sheldrick GM: **How to find more seleniums.** *Abstract PT18 of the Proceedings of the American Crystallographic Association Meeting, May 24–27 1999, Buffalo*.

30. Dauter Z, Dauter M, De La Fortelle E, Bricogne G, Sheldrick GM: **Can anomalous signal of sulfur become a tool for solving protein structures? Data collection and model refinement. Anomalous scatterers in the structure.** *J Mol Biol* 1999, **289**:83-92.

31. Kleywegt GJ, Jones TA: **Template convolution to enhance or detect structural features in macromolecular electron density maps.** *Acta Crystallogr D* 1997, **53**:179-185.

32. Schäfer M, Schneider TR, Sheldrick GM: **Crystal structure of vancomycin.** *Structure* 1996, **4**:1509-1515.

33. Loll PJ, Bevivino AE, Korty BD, Axelsen PH: **Simultaneous recognition of a carboxylate-containing ligand and an intramolecular surrogate ligand in the crystal structure of an asymmetric vancomycin dimer.** *J Am Chem Soc* 1997, **119**:1516-1522.

34. Schäfer M, Sheldrick GM, Bahner I, Lackner H: **Crystal structures of actinomycin D and Z3.** *Angew Chem Int Ed* 1998, **37**:2381-2384.

35. Anderson DH, Weiss MS, Eisenberg D: **A challenging case for protein crystal structure determination: the mating pheromone Er-1 from *Euplotes raikovi*.** *Acta Crystallogr D* 1996, **52**:469-480.

36. Weeks CM, Hauptman HA, Smith GD, Blessing RH, Teeter MM, Miller R: **Crambin: a direct solution for a 400 atom structure.** *Acta Crystallogr D* 1995, **51**:33-38.

37. Aree T, Usón I, Schulz B, Reck G, Hoier H, Sheldrick GM, Saenger W: **Variation of a theme: crystal structure with four octakis (2,3,6-tri-O-methyl)-gamma-cyclodextrin molecules hydrated differently by a total of 19.3 water.** *J Am Chem Soc* 1999, **121**:3321-3327.

38. Prive G, Ogihara N, Wesson L, Cascio D, Eisenberg D: **A designer peptide at high resolution: Shake-and-Bake solution of a 400 atom structure.** *Abstract W008 of the Proceedings of the American Crystallographic Association Meeting, July 23–28 1995, Montreal.*

39. Loll PJ, Miller R, Weeks CM, Axelsen PH: **A ligand-mediated dimerization mode for vancomycin.** *Chem Biol* 1998, **5**:293-298.

40. Smith GD, Blessing RH, Ealick SE, Fontecilla-Camps JC, Hauptman HA, Housset D, Langs DA, Miller R: **The *ab initio* structure determination and refinement of a scorpion protein toxin.** *Acta Crystallogr D* 1997, **53**:551-557.

41. Gessler K, Usón I, Takaha T, Krauss N, Smith SM, Okada S, Sheldrick GM, Saenger W: **V-Amylose at atomic resolution: X-ray structure of a cycloamylose with 26 glucoses.** *Proc Natl Acad Sci USA* 1999, **96**:4246-4251.

42. Kärcher J, Lubini P, Pohl E, Schneider TR, Sheldrick GM: ***Ab initio* solution and refinement of the antibiotic mersacidin, a twinned structure with 850+ atoms.** *Abstract PT17 of the Proceedings of the American Crystallographic Association Meeting, May 22–27 1999, Buffalo*.

43. Parisini E, Capozzi F, Lubini P, Lamzin VS, Luchinat C, Sheldrick GM: ***Ab initio* solution and refinement of two high potential iron protein structures at atomic resolution.** *Acta Crystallogr D* 1999, in press.